Course : Big Data: Practical methods and solutions for data analysis

Practical course - 5d - 35h00 - Ref. BID
Price : 3660 CHF E.T.

★★★★⯪ 4,4 / 5

BEST

Teaching objectives

At the end of the training, the participant will be able to:

- ✓ Understand the concepts and benefits of Big Data with respect to business challenges
- ✓ Understand the technological ecosystem needed to carry out a Big Data project
- ✓ Acquire the technical skills to manage massive, unstructured, complex data flows
- ✓ Implement statistical analysis models to address business needs
- ✓ Learn about a data visualization tool for reporting dynamic analyses

Practical details

Hands-on work

Set up a Hadoop platform and its basic components, use an ETL to manage the data, create analysis modules and dashboards.

Course schedule

1 Understanding the concepts and challenges of Big Data

- Origins and definition of Big Data.
- Key figures in the international and French markets.
- The challenges of Big Data: ROI, organization, data privacy.
- An example of Big Data architecture.

(2) Big Data technologies

- Description of the architecture and components of the Hadoop platform.
- Storage methods (NoSQL, HDFS).
- Operating principles of MapReduce, Spark, Storm, etc.
- Most popular distributions on the market (Hortonworks, Cloudera, MapR, Elastic Map Reduce, Biginsights).
- Installing a Hadoop platform.
- Technologies for the data scientist.

Exercise
Exercise

(3) Installing a Hadoop Big Data platform (via Cloudera Quickstart or other software).

- Operating principles of the Hadoop Distributed File System (HDFS).
- Importing outside data into HDFS.
- Creating SQL requests with HIVE.
- Using PIG to process the data.
- Using an ETL to industrialize the creation of massive data flows.
- Overview of Talend For Big Data.

Exercise
Operating principles of the Hadoop Distributed File System (HDFS).

(4) Importing outside data into HDFS.

- Creating SQL requests with HIVE.
- Using PIG to process the data.
- The principle of ETL (Talend, etc.).
- Managing massive data streaming (NIFI, Kafka, Spark, Storm, etc.)

Exercise
Implementing massive data flows

(5) Big Data Analytics techniques and methods

- Machine Learning: A component of artificial intelligence.
- Discovering the three families: Regression, Classification, and Clustering.
- Data preparation, feature engineering.
- Generating models in R or Python.
- Ensemble Learning.

Exercise
Exercise

(6) Setting up analyses with the tools studied.

- Takeaways.
- Summary of best practices.
- Bibliography.

Dates and locations

REMOTE CLASS

2026 : 18 May, 18 May, 20 July, 20 July, 21 Sep., 16 Nov., 16 Nov.